

# A2 - Estatística Bayesiana

A weakly informative default prior distribution for logistic and other regression models

Professor: Luiz Max Carvalho  
Aluno: Ezequiel de Braga Santos

Julho 2024

- (a) Separação e não identificabilidade são problemas comuns em regressão logística. A primeira ocorre frequentemente em problemas aplicados, onde a variável resposta é perfeitamente predita por uma combinação linear de covariáveis. Nesse contexto, a abordagem Bayesiana surge como solução, inicialmente sendo proposto o uso da priori de Jeffreys, mas que esbarra em problemas de computação e interpretação. Assim, é proposto em [1] uma priori que produz estimativas estáveis e regularizadas, usando uma adaptação de mínimos quadrados ponderados com a família de distribuição  $t$  como priori. Outro problema comum é a esparsidade, quando a dimensão das features é muito alta comparada as amostras de treinamento. Nesse sentido, a priori proposta adiciona informação o suficiente para regularizar inferências extremas obtidas usando máxima verossimilhança ou prioris completamente não informativas, uma vez que prioris fracamente informativas encolhem as estimativas em direção a média quando os dados são esparsos, mas também exercem pouca influência com dados satisfatórios.
- (b) Um desafio ao elicitar uma priori padrão é acertar a escala, por isso há a necessidade de padronizar as variáveis para evitar a preocupação com coeficientes de valores muito grandes ou muito pequenos. O procedimento adotado consiste em fazer entradas binárias terem média 0, e as outras entradas são escalonadas para terem média 0 e desvio-padrão 0,5. Uma consequência disso é que restrições mais rigorosas são aplicadas aos preditores nas interações. Por exemplo, considere três variáveis simétricas  $x_1, x_2$  e  $x_3$ . Ao reescalonar, cada uma delas assumirá os valores  $\pm 1/2$ , o que resulta no ganho de  $\pm 1/4$  numa interação dupla e  $\pm 1/8$  numa interação tripla, o que representa apenas 1/4 do efeito principal, já que todos os coeficientes possuem as mesmas prioris. Isso significa que um salto alto de probabilidade na escala logit é improvável.
- (c) No algoritmo clássico de maximização da verossimilhança, é considerada uma aproximação linearizada da derivada da log-verossimilhança, usando um processo iterativo. Em cada passo, é calculado um pseudo-dado  $z_i$  e uma pseudo-variância  $(\sigma_i^z)^2$ :

$$z_i = X_i \hat{\beta} + \frac{\left(1 + \exp(X_i \hat{\beta})\right)^2}{\exp(X_i \hat{\beta})} \left( y_i - \frac{\exp(X_i \hat{\beta})}{1 + \exp(X_i \hat{\beta})} \right),$$
$$(\sigma_i^z)^2 = \frac{1}{n_i} \frac{\left(1 + \exp(X_i \hat{\beta})\right)^2}{\exp(X_i \hat{\beta})}.$$

A partir disso, aplica-se mínimos quadrados ponderados, com os pesos definidos por  $(\sigma_i^z)^{-2}$ , conforme procedimento geral descrito em [2]. Ao usar prioris  $\beta_j \sim N(\mu_j, \sigma_j^2)$ , é construído um algoritmo para estimar a moda a posteriori  $\hat{\beta}$  usando pseudo-dados  $z_* = (z, \mu)$ ,  $X_* = (X, I)$  e  $w_* = (\sigma^z, \sigma)^{-2}$ , conforme [3].

Por fim o algoritmo apresentado no artigo usa uma adaptação desse algoritmo junto com EM: a ideia é aproximar as distribuições a priori por mistura de gaussianas ( $\beta_j \sim N(\mu_j, \sigma_j^2)$ ,  $\sigma_j \sim \text{Inv} - \chi^2(\nu_j, s_j^2)$ ) e alternar passos de mínimos quadrados ponderados com algoritmo EM, tratando a média dos coeficientes como dados faltantes e estimando  $\sigma_j$ . Assim, o algoritmo pode ser descrito como:

- (i) Dado um  $\beta$ , calcula-se a aproximação normal da verossimilhança e determina  $z$  e  $\sigma^z$  (do algoritmo clássico);
  - (ii) Passo E: aplica uma iteração de mínimos quadrados ponderados e encontra  $\hat{\beta}$  com variância  $V_\beta$ . Calcula o valor esperado da log-posteriori como  $\mathbb{E}[(\beta_j - \mu_j)^2 \mid \sigma, y] \approx (\hat{\beta}_j - \mu_j)^2 + (V_\beta)_{jj}$ ;
  - (iii) Passo M: maximiza o valor esperado anterior e obtém  $\hat{\sigma}_j^2 = \frac{(\hat{\beta}_j - \mu_j)^2 + (V_\beta)_{jj} + \nu_j s_j^2}{1 + \nu_j}$ ;
  - (iv) Recalcula as derivadas usando os pseudo-dados e repete os passos anteriores.
- (d) Para essa análise foram usados os dados do *American National Election Studies* [4] e dados de um pequeno experimento biológico [1]. Para o primeiro, foram consideradas como covariáveis a renda, raça e gênero, com a variável resposta sendo a preferência por candidato republicano ou democrata. Além disso, foram considerados apenas os anos 1988, 1992, 1996 e 2000, após filtragem de dados faltantes/indefinidos. Para o segundo foi considerada o log da dose como covariável e a proporção de mortes como variável resposta. Para ambos, os preditores binários foram padronizados subtraindo a média e o restante, normalizados subtraindo a média e dividindo por duas vezes o desvio-padrão. Para cada um dos dados, foram ajustados modelos com prioris normal, student e cauchy, considerando as funções de ligação logit e probit. Para o ajuste, foi criado um modelo stan. A seguir seguem os resultados da média da distribuição preditiva a priori.

# ANES

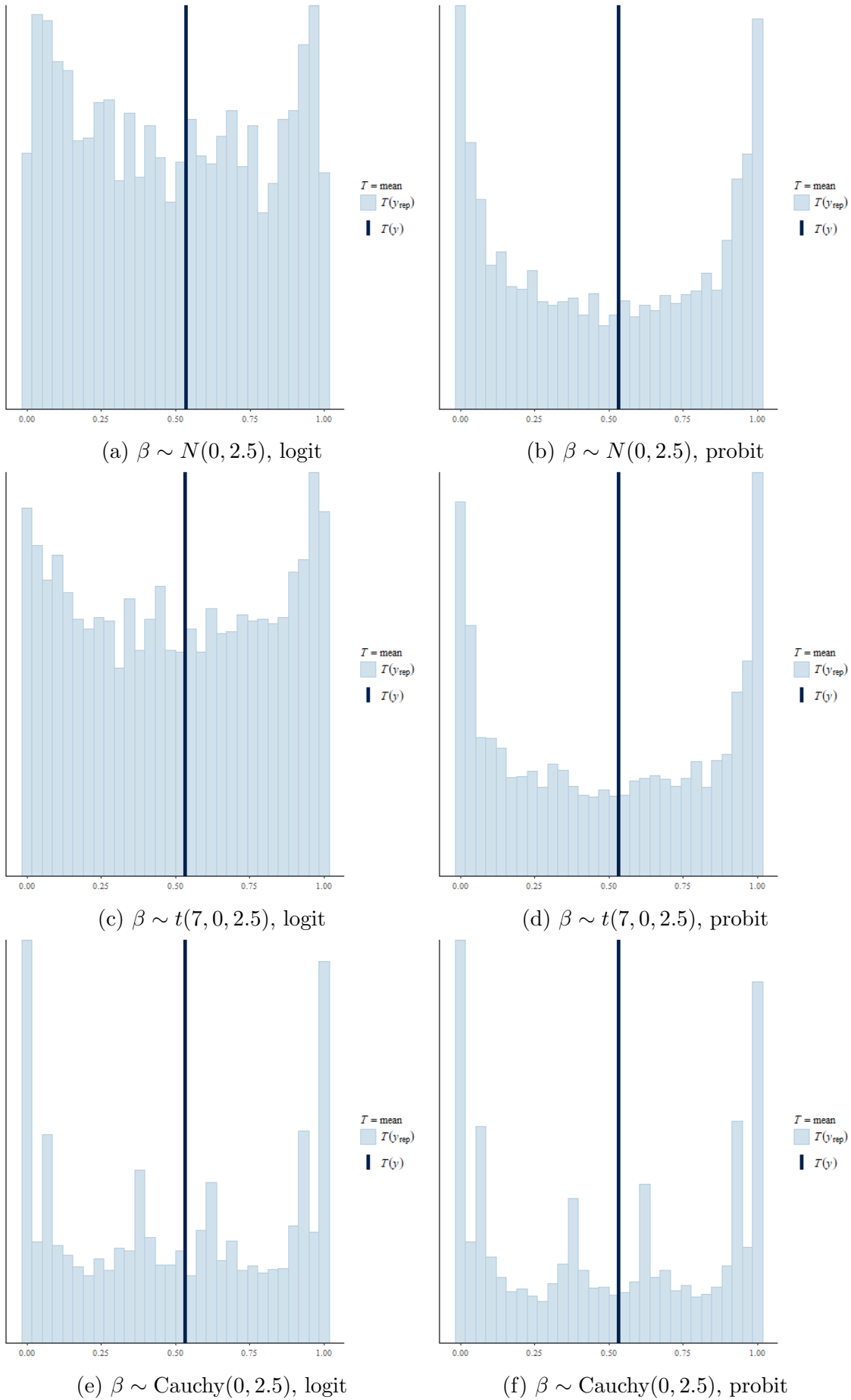
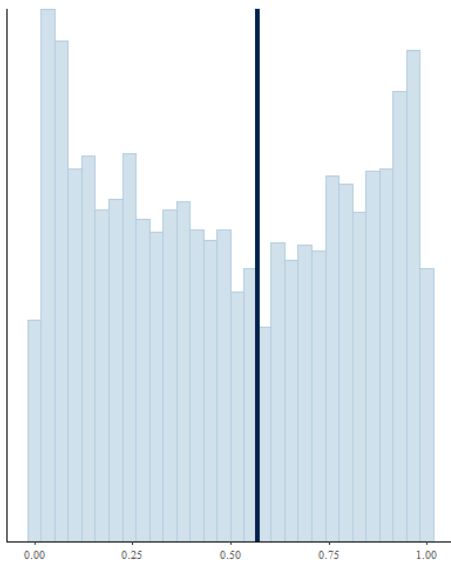
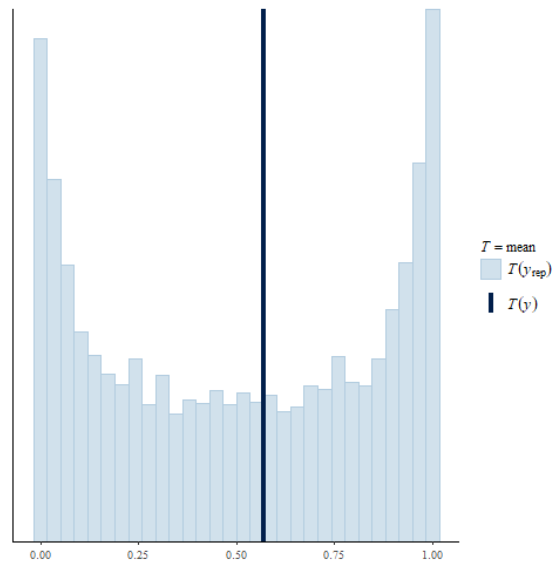


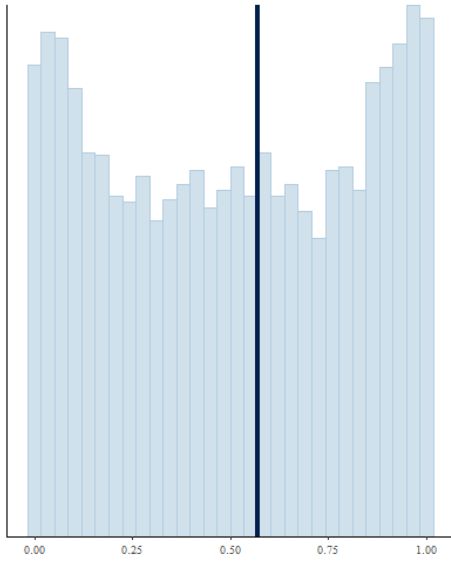
Figura 1: Média preditiva a priori - ANES (1988).



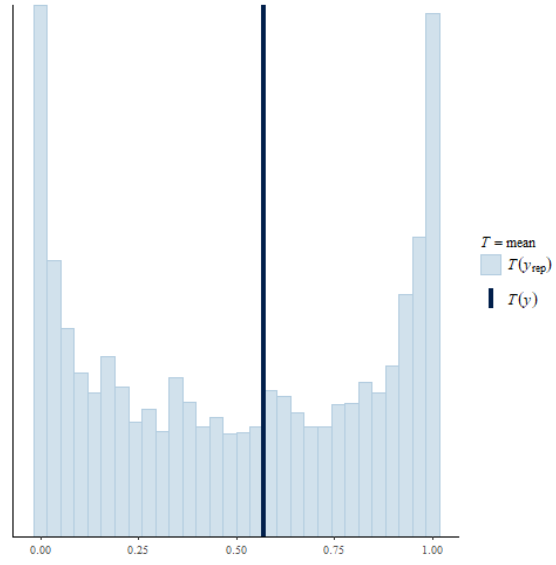
(a)  $\beta \sim N(0, 2.5)$ , logit



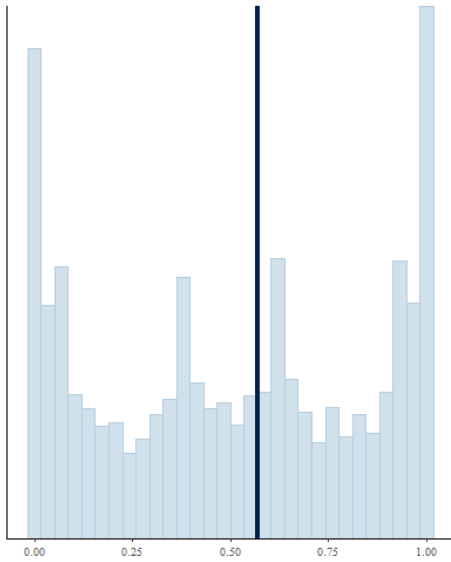
(b)  $\beta \sim N(0, 2.5)$ , probit



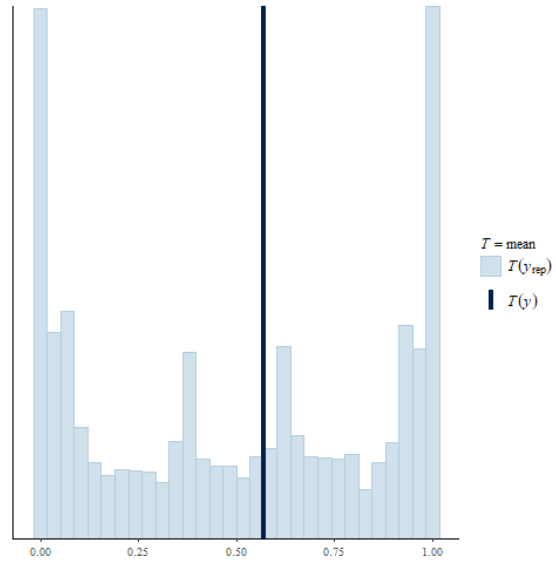
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit

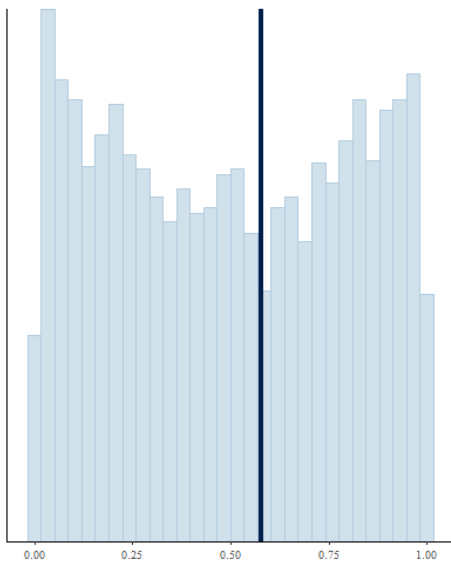


(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit

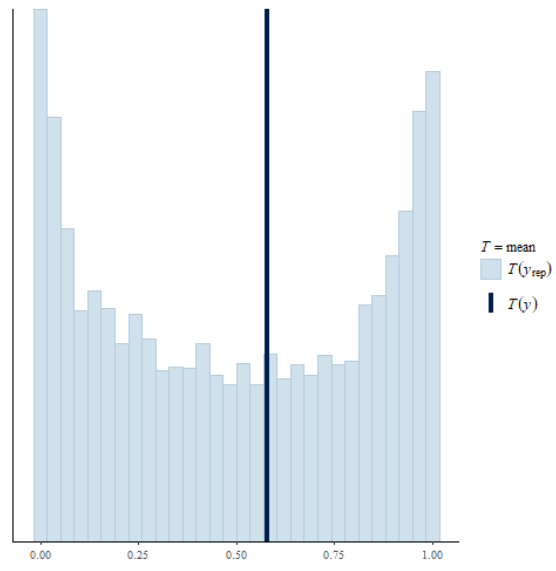


(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

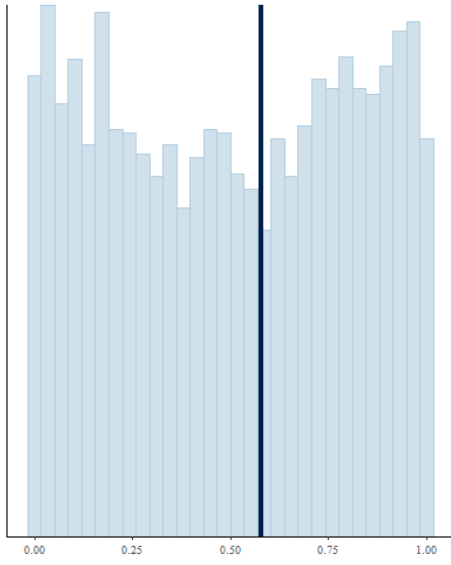
Figura 2: Média preditiva a priori - ANES (1992).



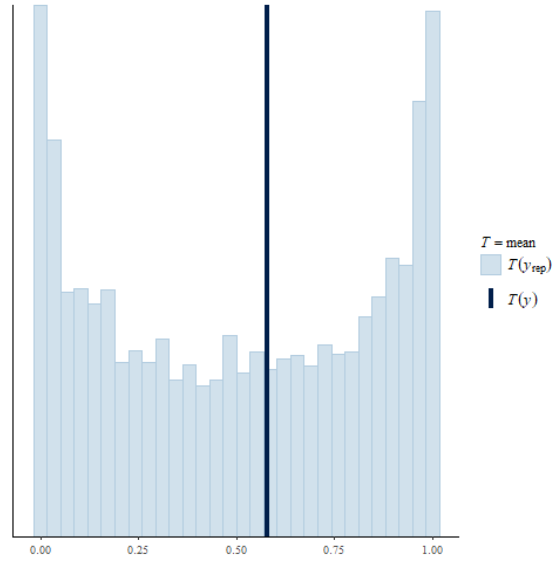
(a)  $\beta \sim N(0, 2.5)$ , logit



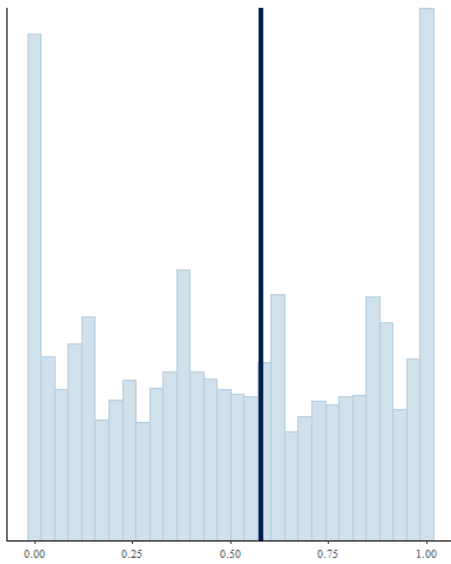
(b)  $\beta \sim N(0, 2.5)$ , probit



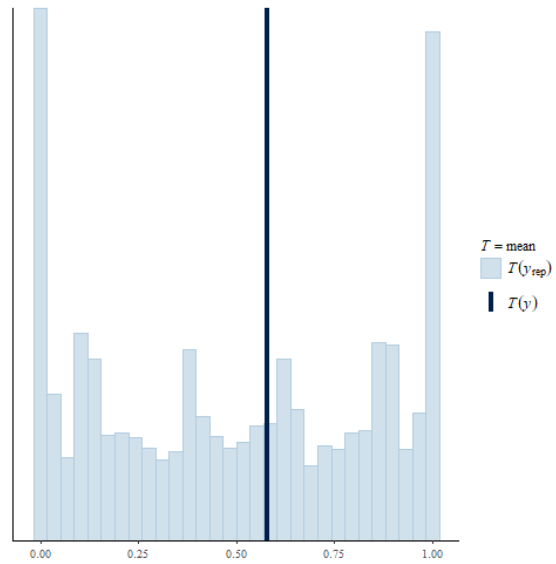
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit

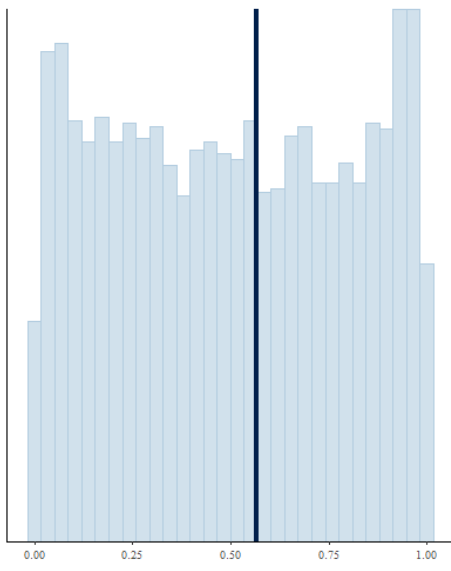


(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit

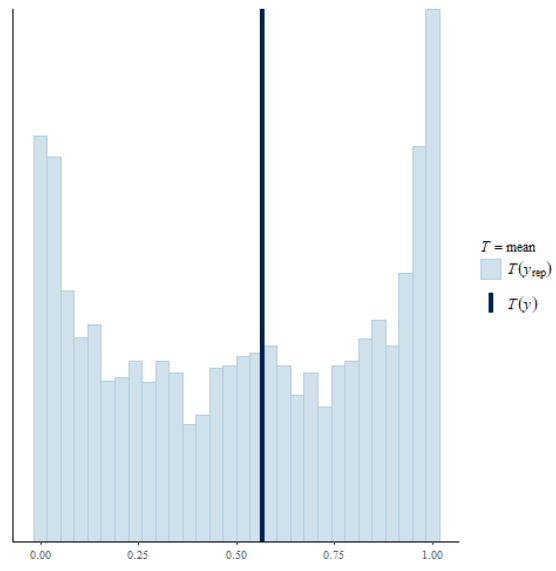


(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

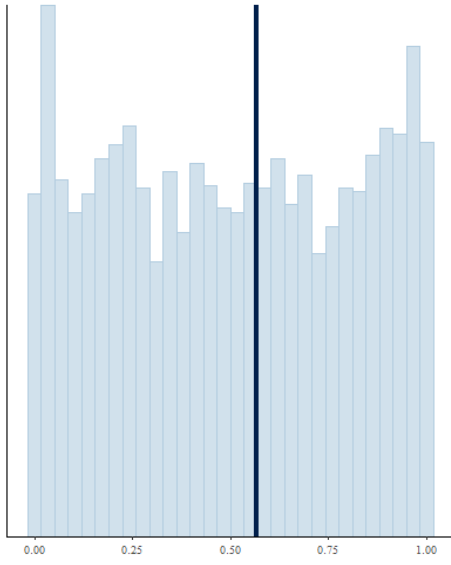
Figura 3: Média preditiva a priori - ANES (1996).



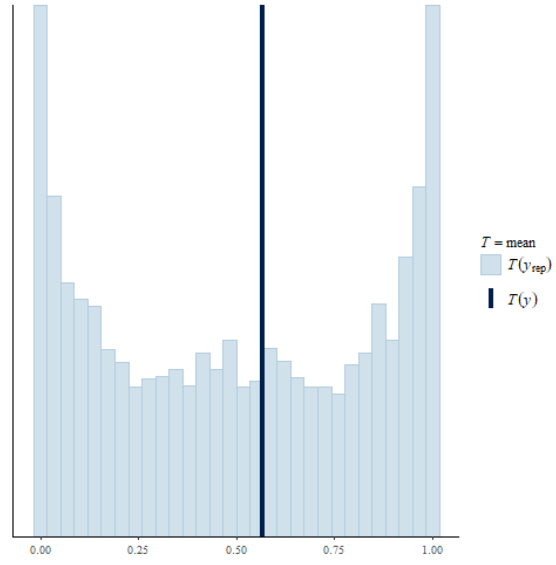
(a)  $\beta \sim N(0, 2.5)$ , logit



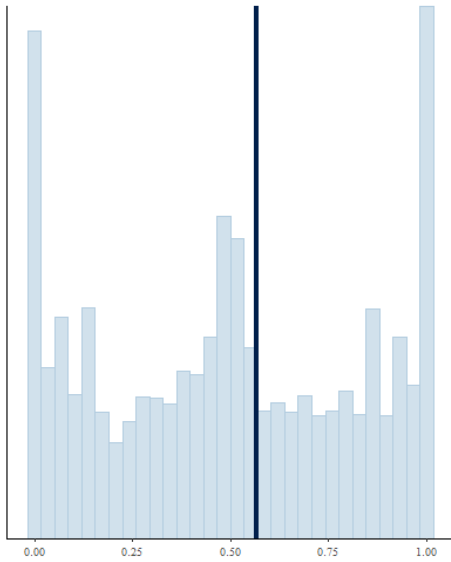
(b)  $\beta \sim N(0, 2.5)$ , probit



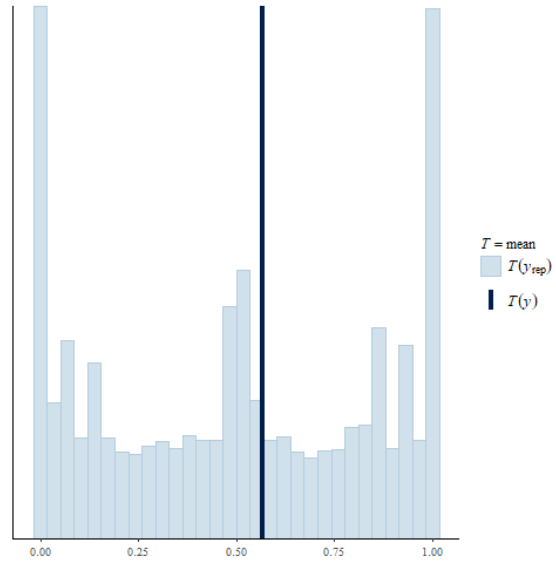
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit



(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit



(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

Figura 4: Média preditiva a priori - ANES (2000).

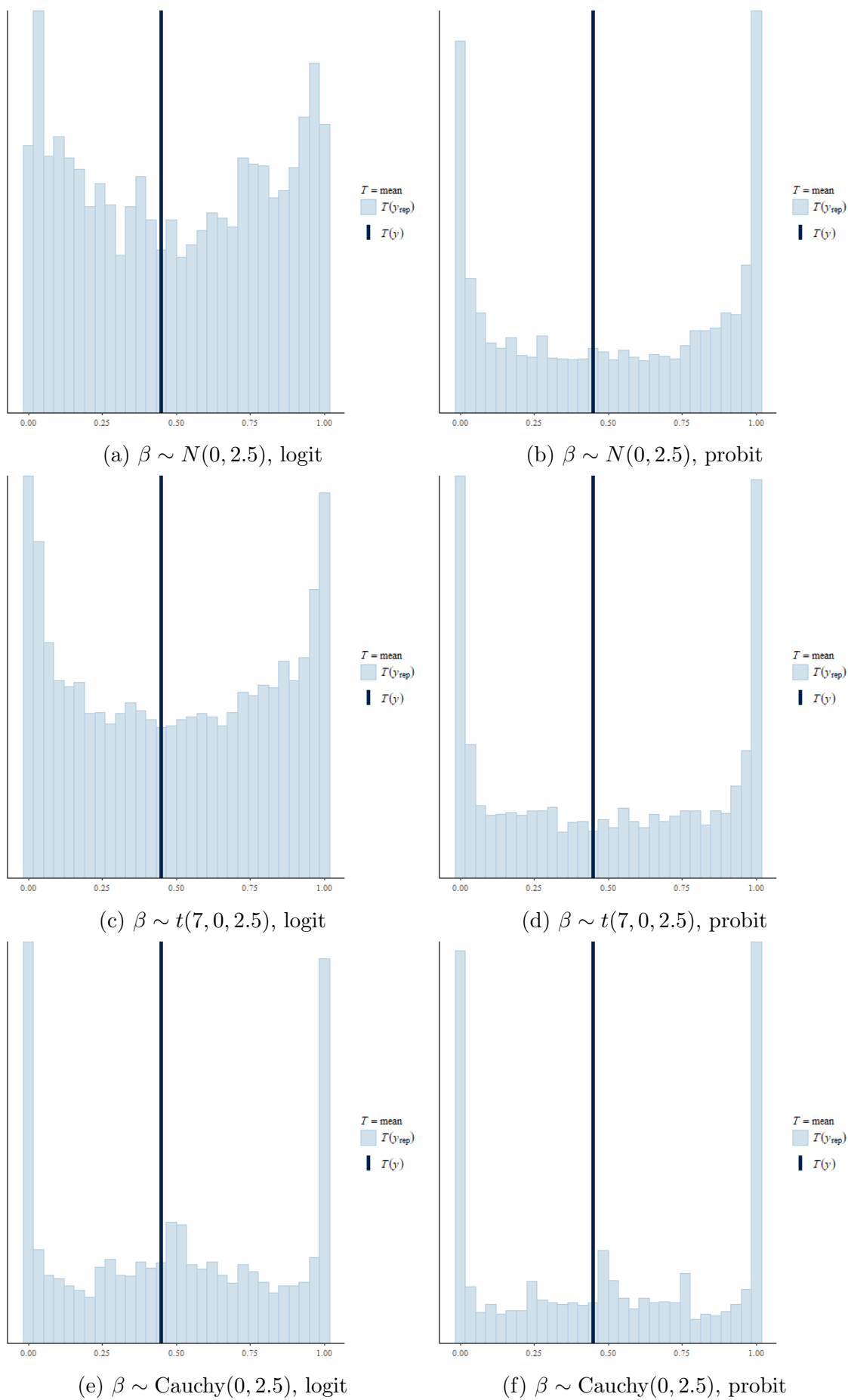


Figura 5: Média preditiva a priori - Bioensaio.

## Comentários

Perceba que todas as prioris colocam massa em todas as regiões plausíveis, que é o desejado. Um comportamento observado é que os modelos probit colocam menos massa em torno da média observada. Além disso, como Cauchy é uma distribuição de cauda pesada, há uma concentração muito menor de probabilidade nessa mesma região quando comparada com normal e student, já que essas distribuições colocam massa de maneira mais “uniforme”.

Observação: discuti essa questão com o Isaque até o último dia e não entendemos com confiança como fazer essa verificação em logística. De todo modo, coloquei os resultados que achamos razoáveis.

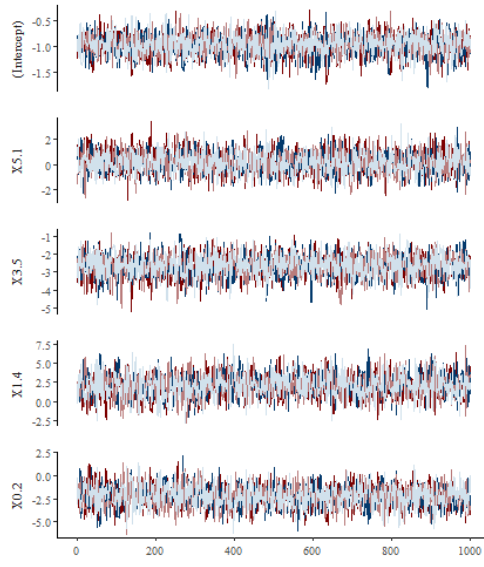
- (e) Para as análises realizadas, foram escolhidos conjuntos de dados pequenos para evitar uma seleção de covariáveis: iris, lenses, hayes, bupa e ring. Todas as covariáveis presentes foram padronizadas (variáveis binárias passaram a ter média 0; variáveis restantes passaram a ter média 0 e desvio-padrão 0.5) e consideradas no ajuste dos modelos. Para cada um dos dados, foram consideradas as distribuições normal, student e cauchy como prioris, com as funções de ligação logit e probit, usando a função *stan\_glm()* do pacote **rstanarm**. Para diagnóstico dos modelos, foram usados mcse, n\_eff, Rhat, trace plots e intervalos de credibilidade. Por fim, para comparar os modelos, foi utilizada validação cruzada, usando a função *loo\_compare()* do pacote **loo**. A seguir estão os resultados.



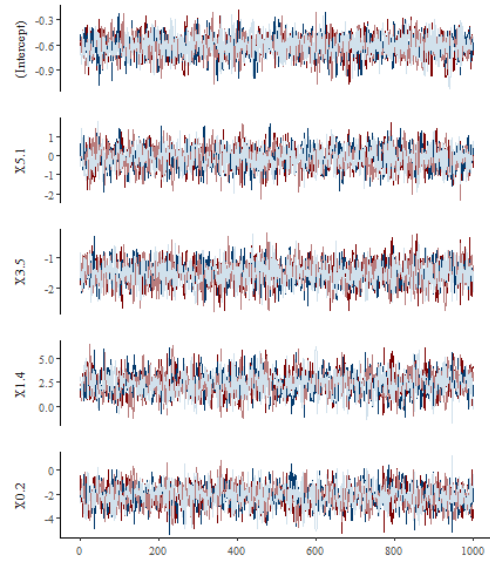
# Iris

Priori	Função de ligação		mcse	Rhat	n_eff
$N(0, 2.5)$	logit	(Intercept)	0	1	2490
		X5.1	0	1	2160
		X3.5	0	1	2325
		X1.4	0	1	1881
		X0.2	0	1	2201
		mean_PPD	0	1	3124
		log-posterior	0	1	1484
$N(0, 2.5)$	probit	(Intercept)	0	1	2304
		X5.1	0	1	1794
		X3.5	0	1	2124
		X1.4	0	1	1433
		X0.2	0	1	1539
		mean_PPD	0	1	3224
		log-posterior	0	1	1523
$t(7, 0, 2.5)$	logit	(Intercept)	0	1	2688
		X5.1	0	1	2255
		X3.5	0	1	2239
		X1.4	0	1	1639
		X0.2	0	1	1985
		mean_PPD	0	1	3461
		log-posterior	0	1	1577
$t(7, 0, 2.5)$	probit	(Intercept)	0	1	2449
		X5.1	0	1	2097
		X3.5	0	1	2107
		X1.4	0	1	1696
		X0.2	0	1	2063
		mean_PPD	0	1	3642
		log-posterior	0	1	1589
Cauchy(0, 2.5)	logit	(Intercept)	0	1	2502
		X5.1	0	1	1491
		X3.5	0	1	1624
		X1.4	0.1	1	1154
		X0.2	0	1	1457
		mean_PPD	0	1	3634
		log-posterior	0	1	1524
Cauchy(0, 2.5)	probit	(Intercept)	0	1	2592
		X5.1	0	1	1819
		X3.5	0	1	1982
		X1.4	0	1	1218
		X0.2	0	1	1424
		mean_PPD	0	1	3269
		log-posterior	0	1	1604

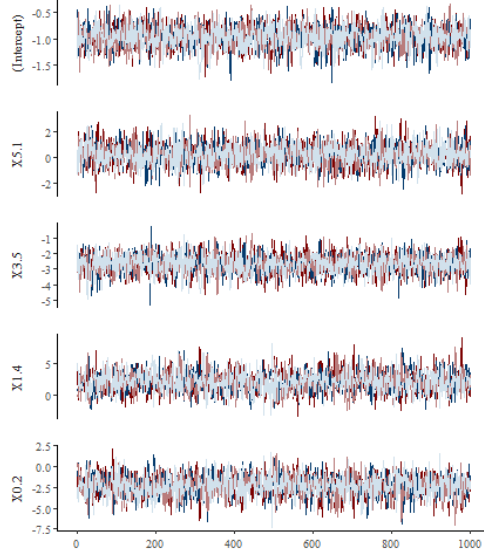
Tabela 1: Resultados - Iris



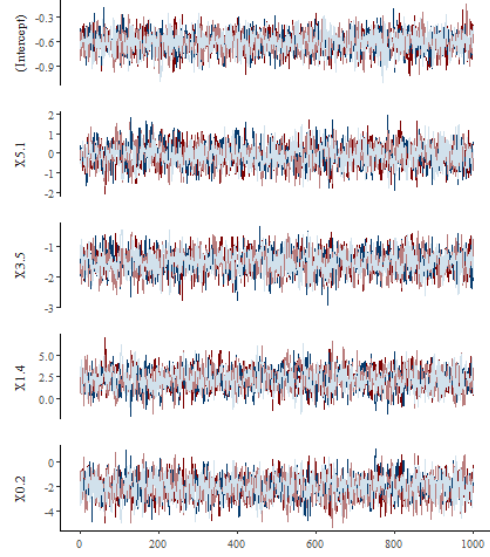
(a)  $\beta \sim N(0, 2.5)$ , logit



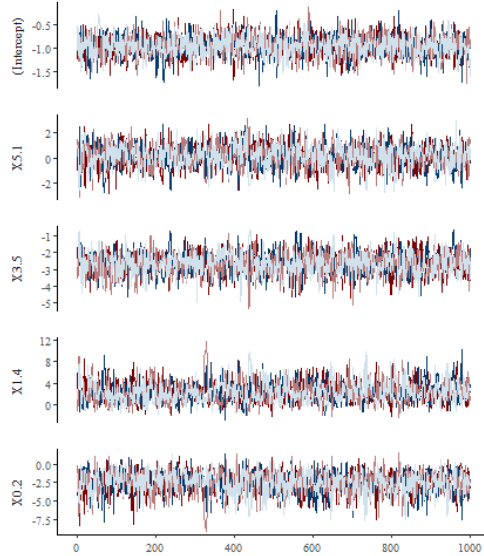
(b)  $\beta \sim N(0, 2.5)$ , probit



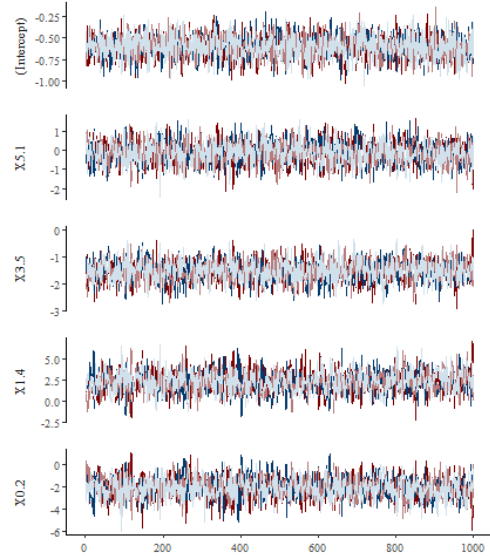
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit

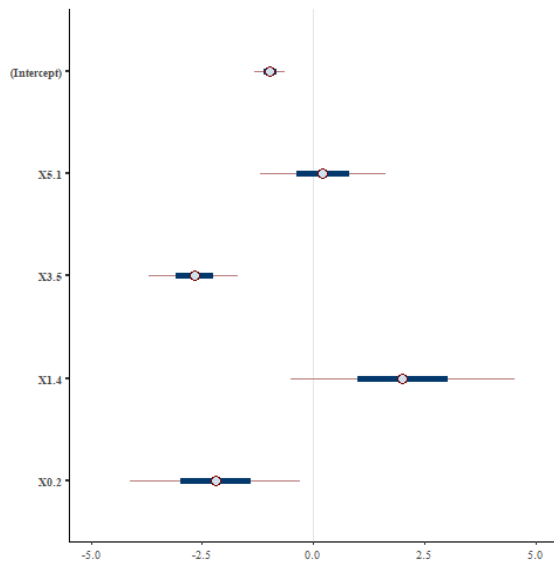


(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit

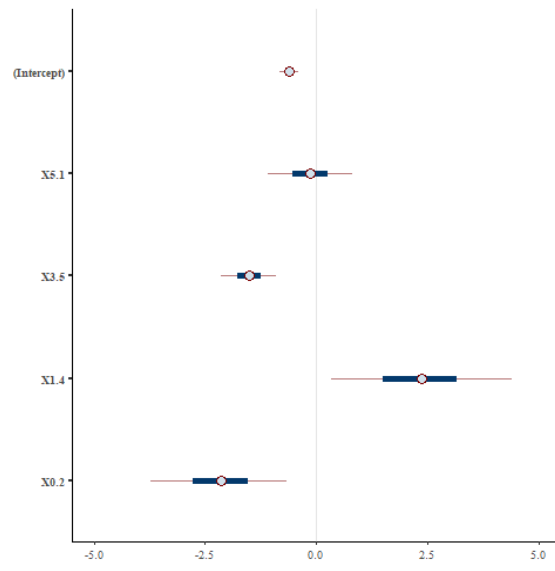


(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

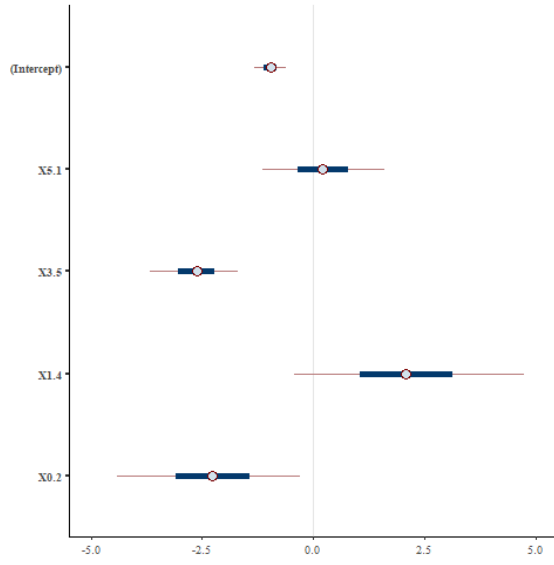
Figure 6: Trace plots - Iris.



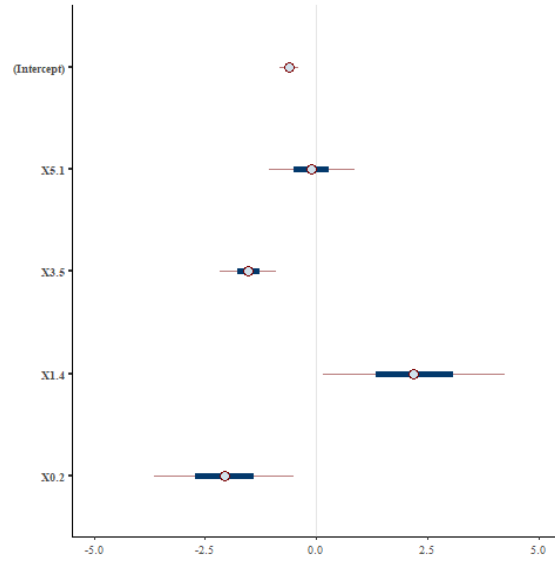
(a)  $\beta \sim N(0, 2.5)$ , logit



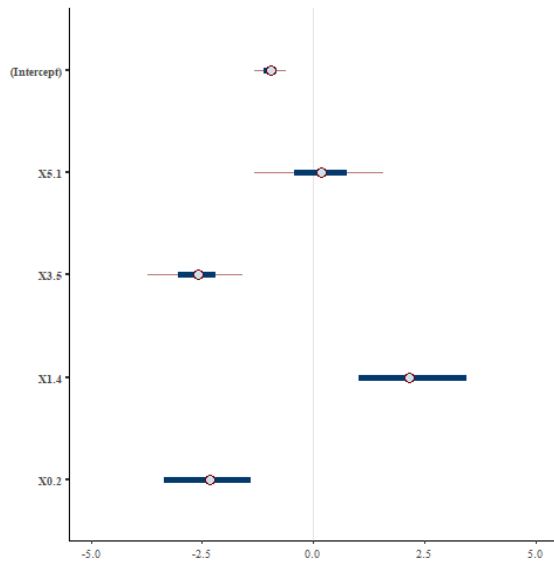
(b)  $\beta \sim N(0, 2.5)$ , probit



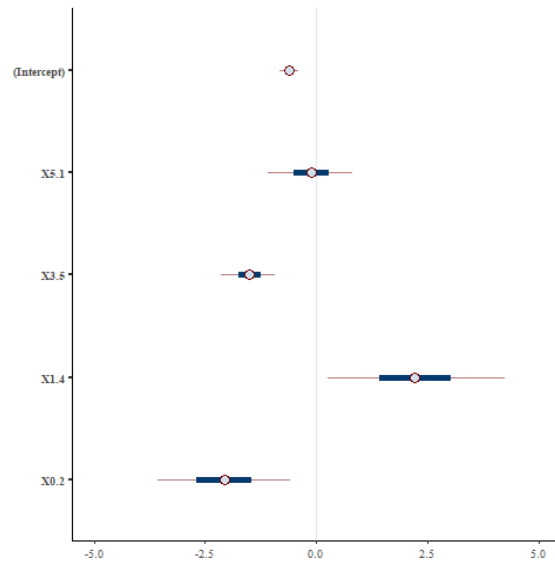
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit



(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit



(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

Figura 7: Intervalos de credibilidade - Iris.

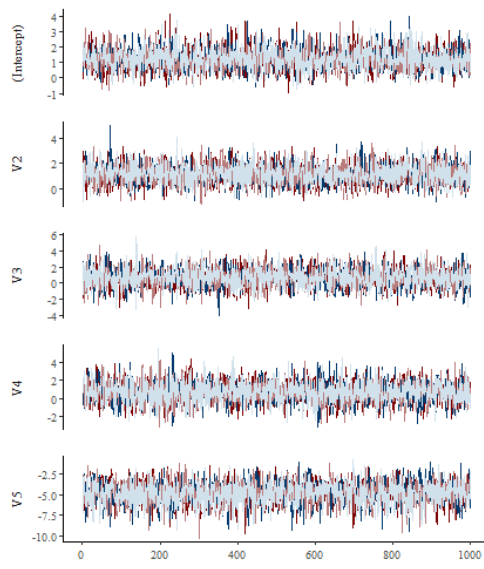
	elp_diff	se_diff
iris_fit_32	0.0	0.0
iris_fit_12	0.0	0.1
iris_fit_22	-0.1	0.1
iris_fit_21	-0.2	0.9
iris_fit_11	-0.2	1.0
iris_fit_31	-0.4	0.8

Tabela 2: Comparação dos modelos usando *loo\_compare()* - Iris.

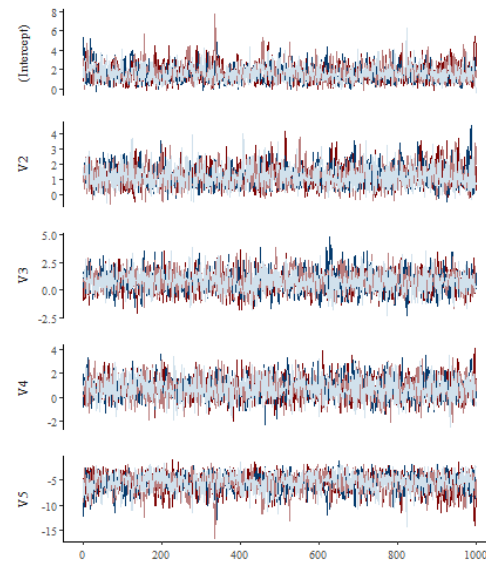
# Lenses

Priori	Função de ligação		mcse	Rhat	n_eff
$N(0, 2.5)$	logit	(Intercept)	0	1	3076
		V2	0	1	3299
		V3	0	1	3940
		V4	0	1	3802
		V5	0	1	2604
		mean_PPD	0	1	4512
		log-posterior	0	1	1595
$N(0, 2.5)$	probit	(Intercept)	0	1	1536
		V2	0	1	2174
		V3	0	1	2673
		V4	0	1	3180
		V5	0	1	1515
		mean_PPD	0	1	4062
		log-posterior	0	1	1510
$t(7, 0, 2.5)$	logit	(Intercept)	0	1	1940
		V2	0	1	3324
		V3	0	1	3355
		V4	0	1	3052
		V5	0	1	1825
		mean_PPD	0	1	4829
		log-posterior	0	1	1831
$t(7, 0, 2.5)$	probit	(Intercept)	0	1	1080
		V2	0	1	2247
		V3	0	1	2157
		V4	0	1	2925
		V5	0.1	1	1062
		mean_PPD	0	1	4052
		log-posterior	0	1	1344
Cauchy(0, 2.5)	logit	(Intercept)	0	1	461
		V2	0	1	2638
		V3	0	1	2790
		V4	0	1	2931
		V5	0.2	1	542
		mean_PPD	0	1	4297
		log-posterior	0	1	1282
Cauchy(0, 0.5)	probit	(Intercept)	0	1	923
		V2	0	1	2318
		V3	0	1	2534
		V4	0	1	2539
		V5	0.1	1	1059
		mean_PPD	0	1	4164
		log-posterior	0	1	1618

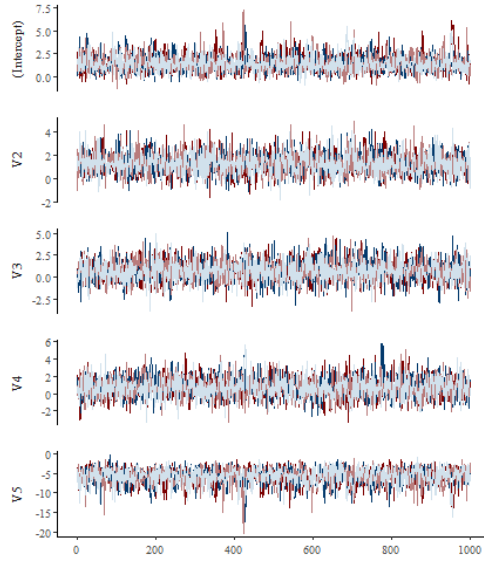
Tabela 3: Resultados - Lenses



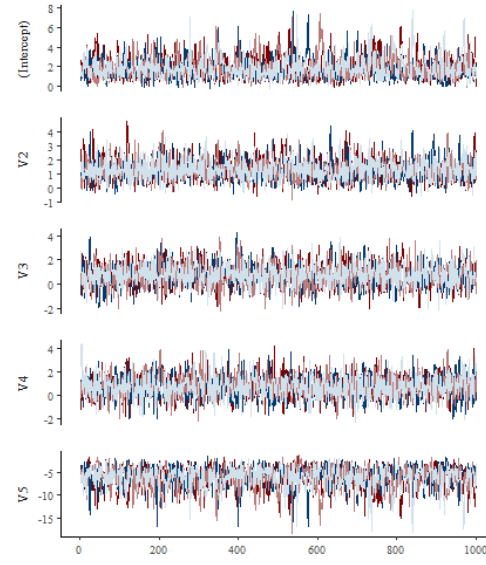
(a)  $\beta \sim N(0, 2.5)$ , logit



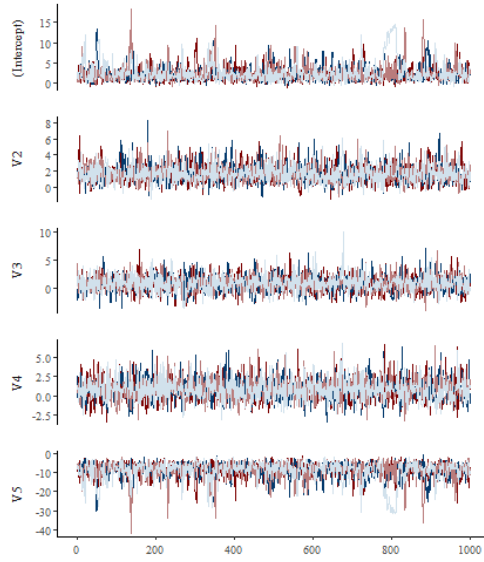
(b)  $\beta \sim N(0, 2.5)$ , probit



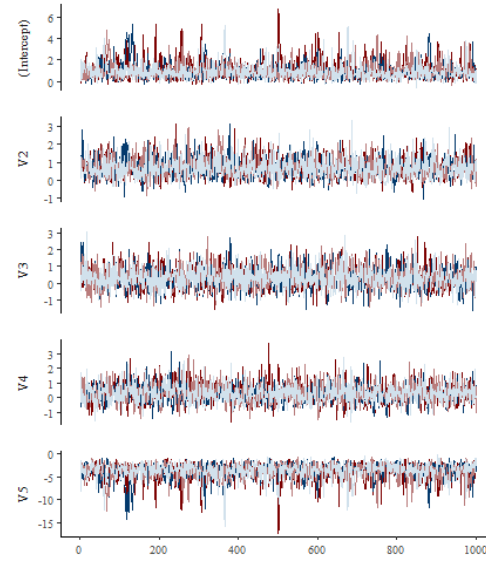
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit



(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit



(f)  $\beta \sim \text{Cauchy}(0, 0.5)$ , probit

Figure 8: Trace plots - Lenses.

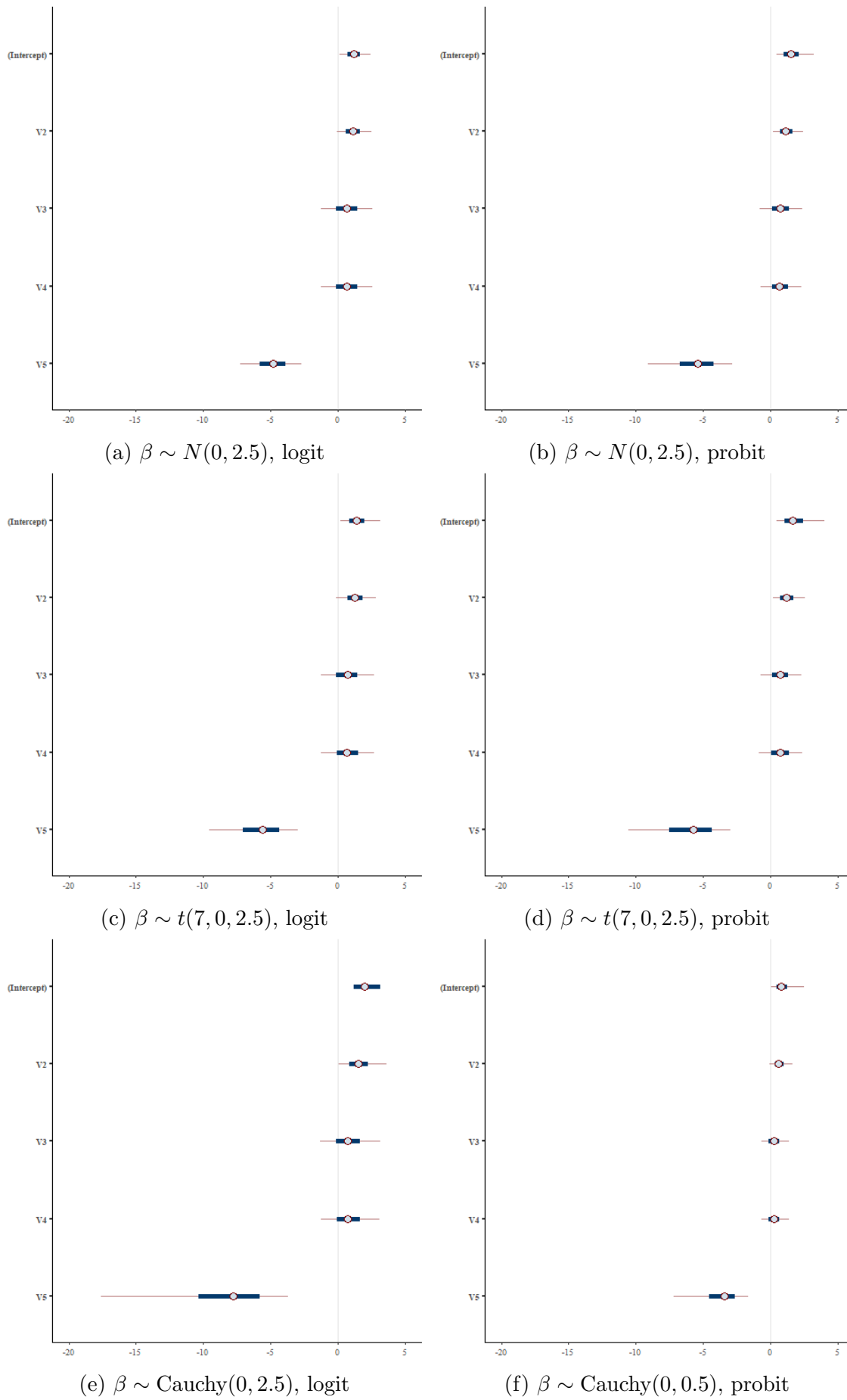


Figura 9: Intervalos de credibilidade - Lentes.

	elp_diff	se_diff
lenses_fit_21	0.0	0.0
lenses_fit_32	0.0	0.7
lenses_fit_11	-0.2	0.3
lenses_fit_31	-0.7	1.1
lenses_fit_12	-1.6	1.9
lenses_fit_22	-2.3	2.2

Tabela 4: Comparação dos modelos usando *loo\_compare()* - Lenses.



# Hayes

Priori	Função de ligação		mcse	Rhat	n_eff
$N(0, 2.5)$	logit	(Intercept)	0	1	3331
		X2	0	1	4687
		X1	0	1	3038
		X1.1	0	1	3539
		X2.1	0	1	3539
		mean_PPD	0	1	5106
		log-posterior	0	1	1698
$N(0, 2.5)$	probit	(Intercept)	0	1	3766
		X2	0	1	4893
		X1	0	1	3453
		X1.1	0	1	3661
		X2.1	0	1	4033
		mean_PPD	0	1	4817
		log-posterior	0	1	1806
$t(7, 0, 2.5)$	logit	(Intercept)	0	1	3353
		X2	0	1	4878
		X1	0	1	3349
		X1.1	0	1	3316
		X2.1	0	1	3819
		mean_PPD	0	1	4420
		log-posterior	0	1	1588
$t(7, 0, 2.5)$	probit	(Intercept)	0	1	3492
		X2	0	1	4530
		X1	0	1	3947
		X1.1	0	1	3459
		X2.1	0	1	3743
		mean_PPD	0	1	4770
		log-posterior	0	1	1704
Cauchy(0, 2.5)	logit	(Intercept)	0	1	3517
		X2	0	1	4536
		X1	0	1	3316
		X1.1	0	1	3750
		X2.1	0	1	3160
		mean_PPD	0	1	3960
		log-posterior	0	1	1895
Cauchy(0, 2.5)	probit	(Intercept)	0	1	3835
		X2	0	1	5214
		X1	0	1	3639
		X1.1	0	1	3655
		X2.1	0	1	4280
		mean_PPD	0	1	4538
		log-posterior	0	1	1826

Tabela 5: Resultados - Hayes

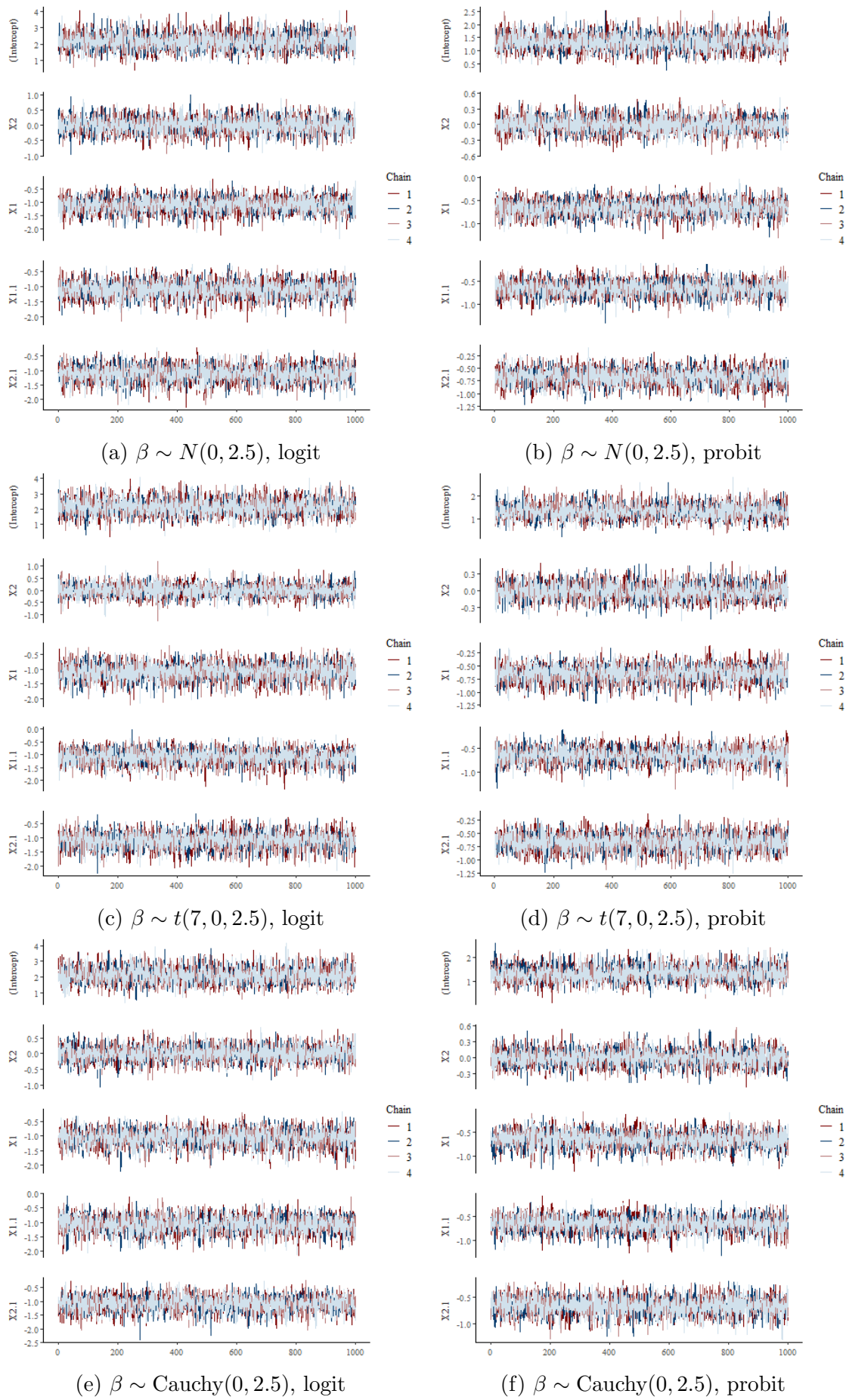
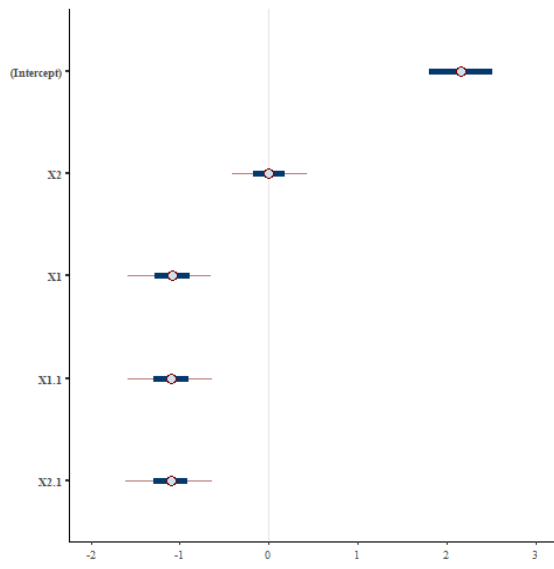
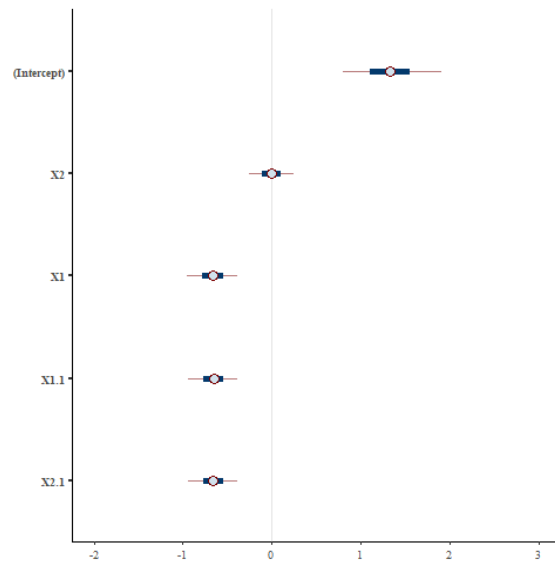


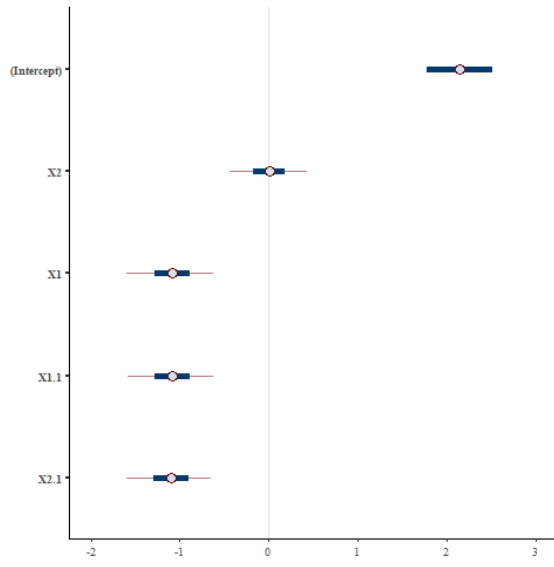
Figure 10: Trace plots - Hayes.



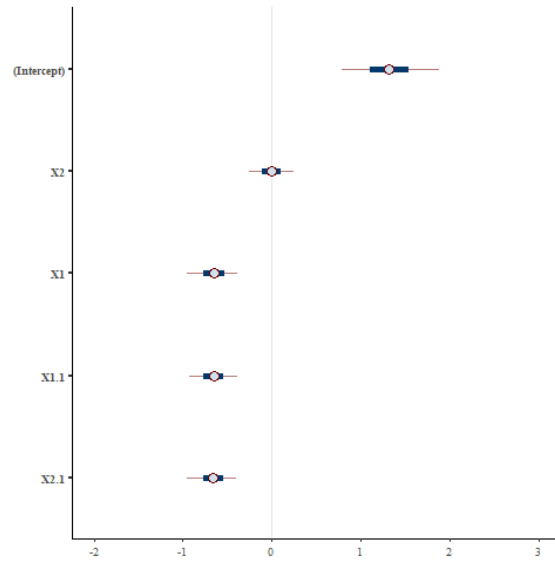
(a)  $\beta \sim N(0, 2.5)$ , logit



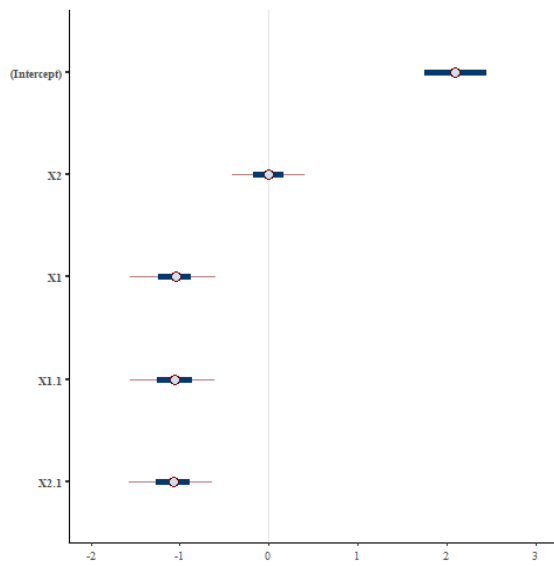
(b)  $\beta \sim N(0, 2.5)$ , probit



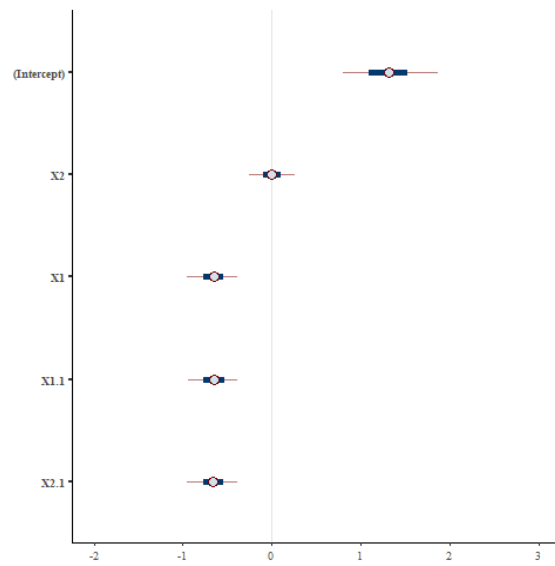
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit



(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit



(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

Figura 11: Intervalos de credibilidade - Hayes.

	elp_diff	se_diff
hayes_fit_22	0.0	0.0
hayes_fit_32	0.1	0.0
hayes_fit_31	-0.1	0.3
hayes_fit_11	-0.2	0.3
hayes_fit_12	-0.2	0.1
hayes_fit_21	-0.2	0.3

Tabela 6: Comparação dos modelos usando *loo\_compare()* - Hayes.

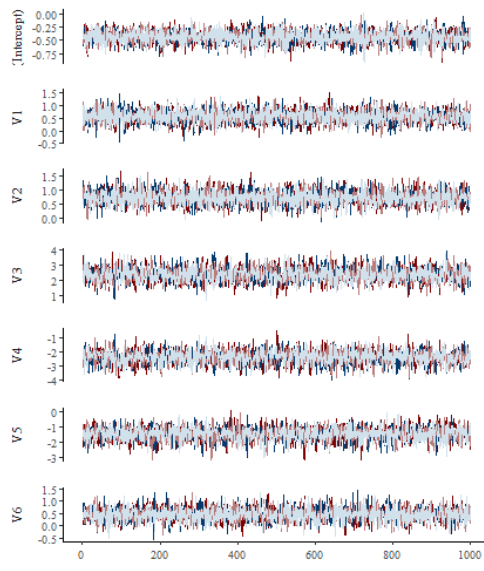
# Bupa

Priori	Função de ligação		mcse	Rhat	n_eff
$N(0, 2.5)$	logit	(Intercept)	0	1	4261
		V1	0	1	4241
		V2	0	1	4173
		V3	0	1	2326
		V4	0	1	2462
		V5	0	1	3093
		V6	0	1	3625
		mean_PPD	0	1	4253
		log-posterior	0	1	1713
$N(0, 2.5)$	probit	(Intercept)	0	1	4805
		V1	0	1	4747
		V2	0	1	4849
		V3	0	1	3298
		V4	0	1	3218
		V5	0	1	4110
		V6	0	1	4408
		mean_PPD	0	1	4096
		log-posterior	0	1	1622
$t(7, 0, 2.5)$	logit	(Intercept)	0	1	4368
		V1	0	1	4567
		V2	0	1	3948
		V3	0	1	3017
		V4	0	1	3289
		V5	0	1	2787
		V6	0	1	4763
		mean_PPD	0	1	4567
		log-posterior	0.1	1	1494
$t(7, 0, 2.5)$	probit	(Intercept)	0	1	4631
		V1	0	1	4063
		V2	0	1	3879
		V3	0	1	2919
		V4	0	1	2990
		V5	0	1	3696
		V6	0	1	4534
		mean_PPD	0	1	4024
		log-posterior	0	1	1881
Cauchy(0, 2.5)	logit	(Intercept)	0	1	3944
		V1	0	1	5021
		V2	0	1	3817
		V3	0	1	2805
		V4	0	1	2907
		V5	0	1	3622
		V6	0	1	3868
		mean_PPD	0	1	4144
		log-posterior	0	1	1787

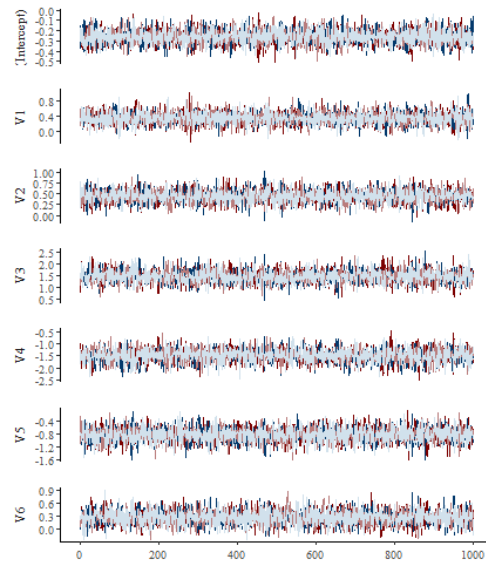
Tabela 7: Resultados - Bupa

Priori	Função de ligação		mcse	Rhat	n_eff
Cauchy(0, 2.5)	probit	(Intercept)	0	1	4597
		V1	0	1	4095
		V2	0	1	4582
		V3	0	1	2765
		V4	0	1	2974
		V5	0	1	4173
		V6	0	1	3542
		mean_PPD	0	1	3993
		log-posterior	0	1	1664

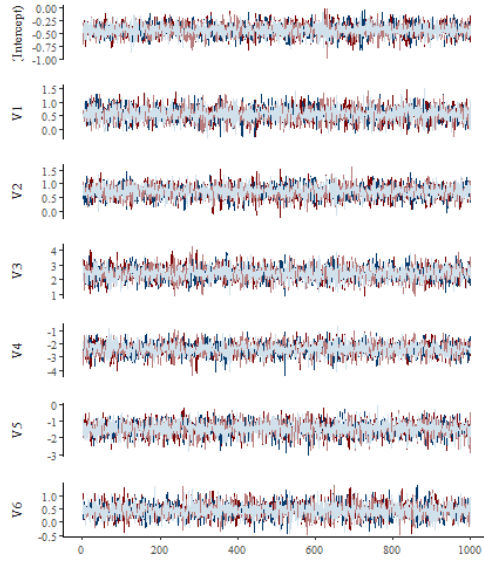
Tabela 8: Resultados - Bupa



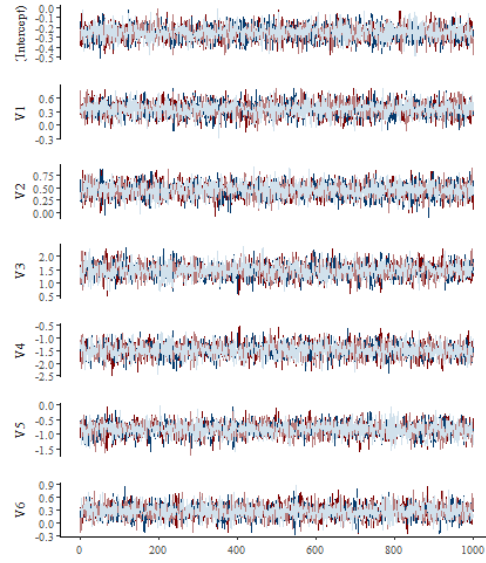
(a)  $\beta \sim N(0, 2.5)$ , logit



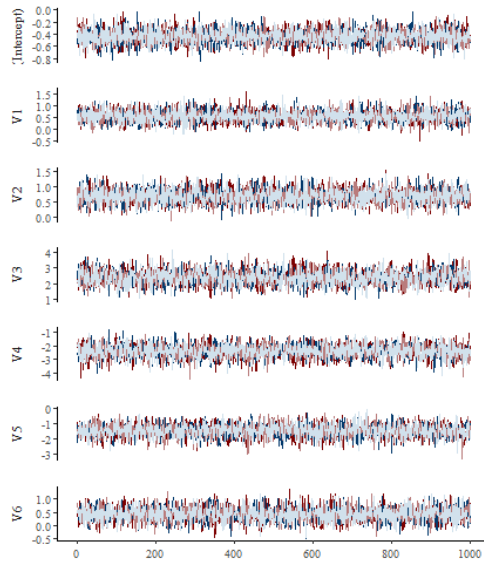
(b)  $\beta \sim N(0, 2.5)$ , probit



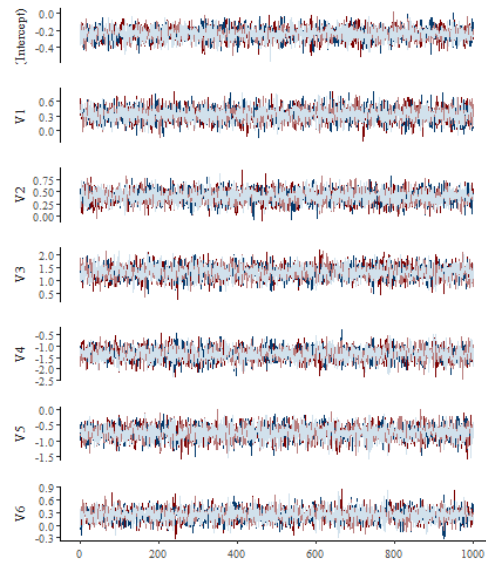
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit

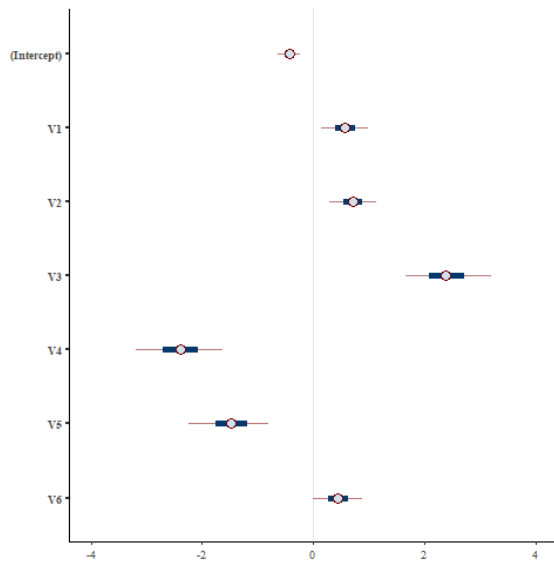


(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit

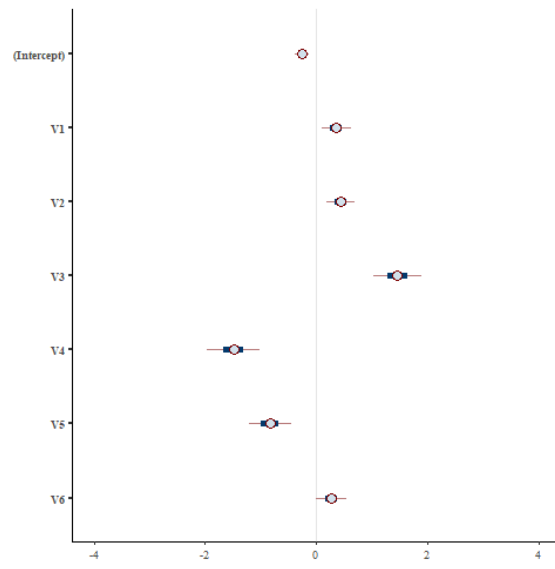


(f)  $\beta \sim \text{Cauchy}(0, 0.5)$ , probit

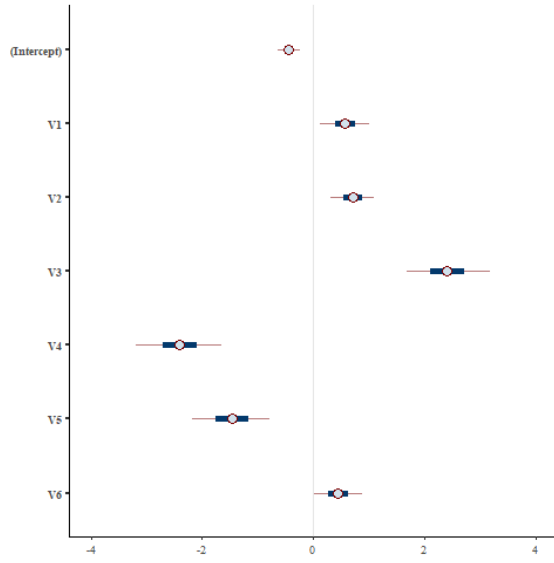
Figure 12: Trace plots - Bupa.



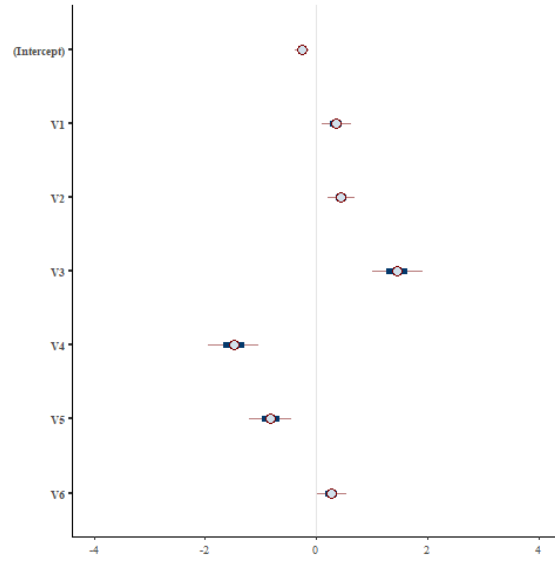
(a)  $\beta \sim N(0, 2.5)$ , logit



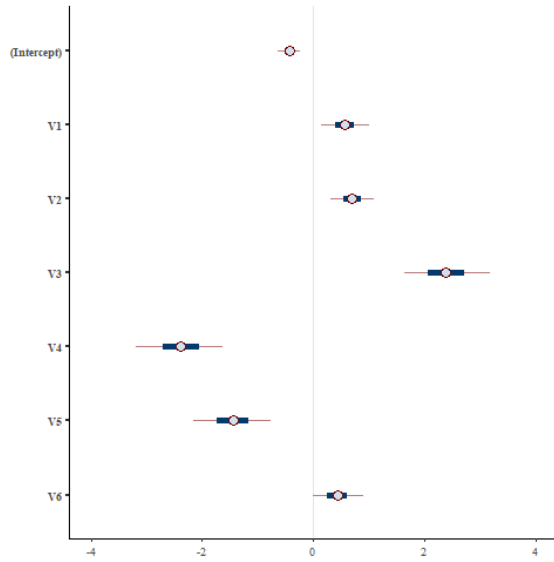
(b)  $\beta \sim N(0, 2.5)$ , probit



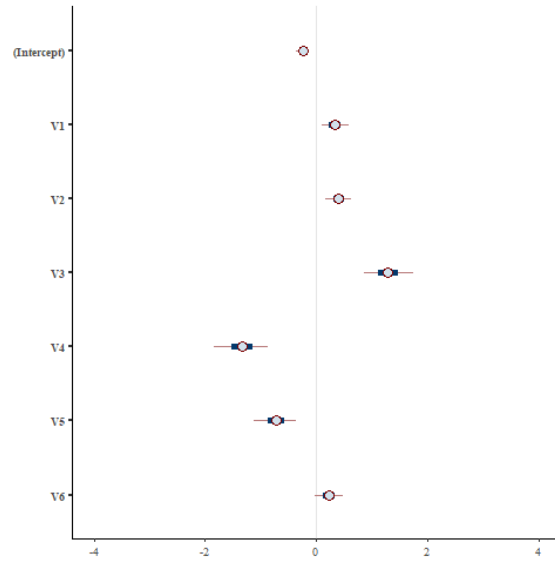
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit



(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit



(f)  $\beta \sim \text{Cauchy}(0, 0.5)$ , probit

Figura 13: Intervalos de credibilidade - Bupa.



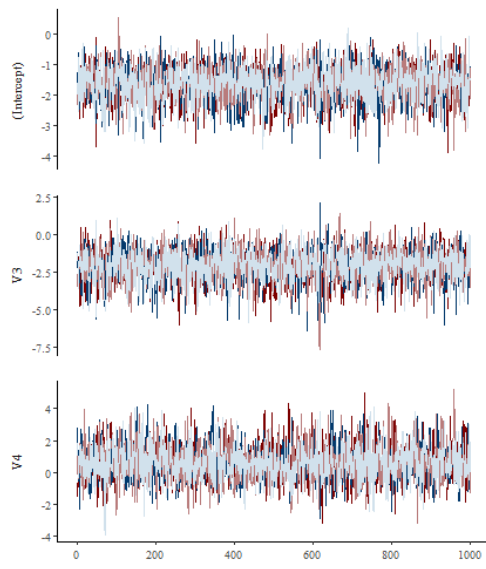
	elp_diff	se_diff
bupa_fit_21	0.0	0.0
bupa_fit_32	-0.2	0.1
bupa_fit_11	-0.3	0.1
bupa_fit_31	-0.6	0.8
bupa_fit_12	-0.7	0.4
bupa_fit_22	-0.8	0.5

Tabela 9: Comparação dos modelos usando *loo\_compare()* - Bupa.

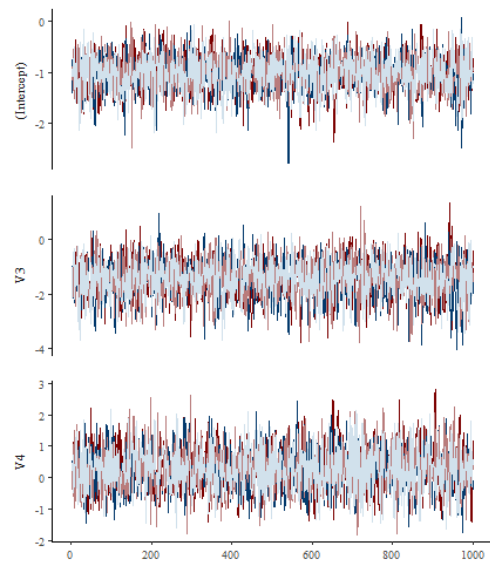
## Ring

Priori	Função de ligação		mcse	Rhat	n_eff
$N(0, 2.5)$	logit	(Intercept)	0	1	2593
		V3	0	1	2781
		V4	0	1	3267
		mean_PPD	0	1	3739
		log-posterior	0	1	1524
$N(0, 2.5)$	probit	(Intercept)	0	1	2859
		V3	0	1	2869
		V4	0	1	3186
		mean_PPD	0	1	3952
		log-posterior	0	1	1951
$t(7, 0, 2.5)$	logit	(Intercept)	0	1	2708
		V3	0	1	2766
		V4	0	1	2866
		mean_PPD	0	1	3967
		log-posterior	0	1	1643
$t(7, 0, 2.5)$	probit	(Intercept)	0	1	2719
		V3	0	1	2605
		V4	0	1	3214
		mean_PPD	0	1	3779
		log-posterior	0	1	1655
Cauchy(0, 2.5)	logit	(Intercept)	0	1	2347
		V3	0	1	2888
		V4	0	1	2771
		mean_PPD	0	1	3774
		log-posterior	0	1	1650
Cauchy(0, 2.5)	probit	(Intercept)	0	1	2612
		V3	0	1	2893
		V4	0	1	2972
		mean_PPD	0	1	3712
		log-posterior	0	1	1625

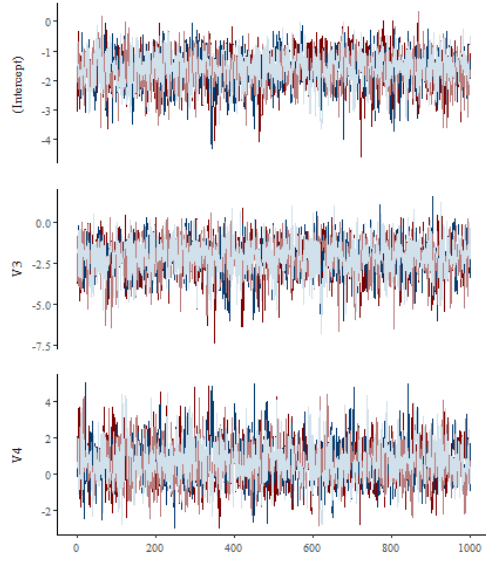
Tabela 10: Resultados - Ring



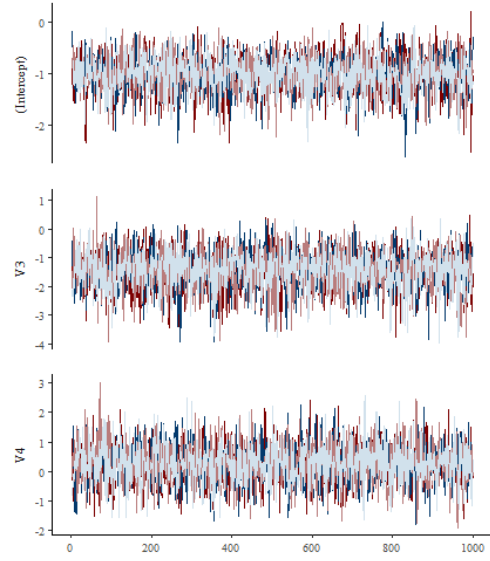
(a)  $\beta \sim N(0, 2.5)$ , logit



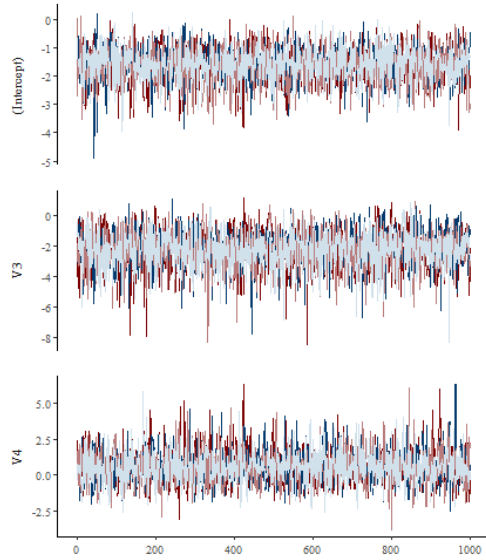
(b)  $\beta \sim N(0, 2.5)$ , probit



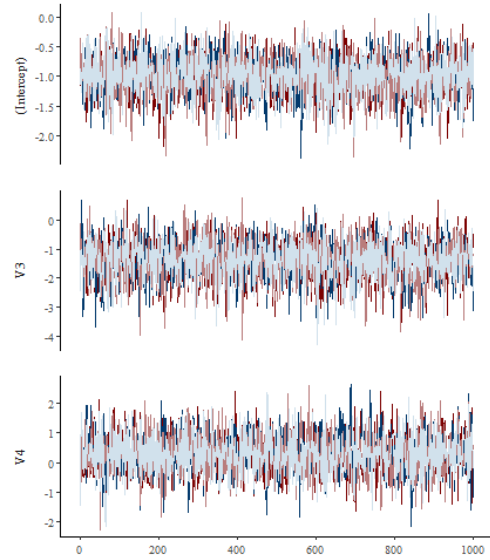
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit

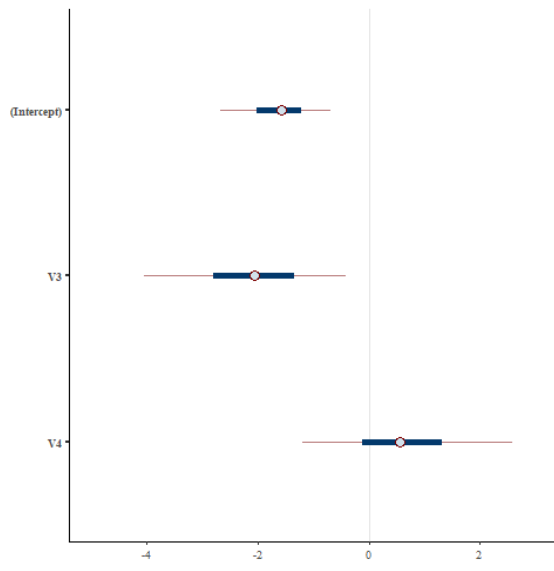


(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit

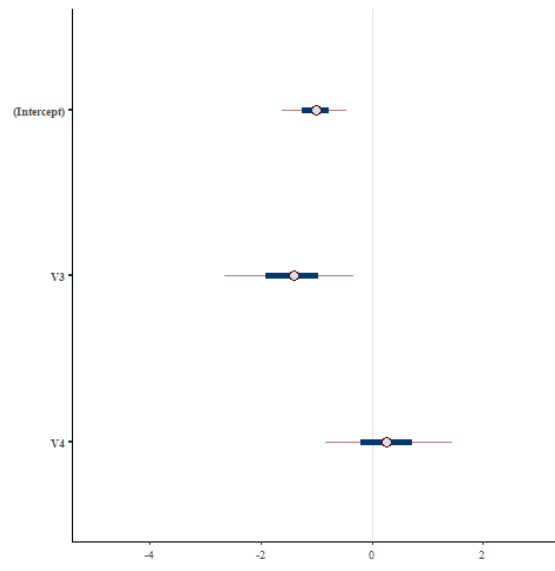


(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

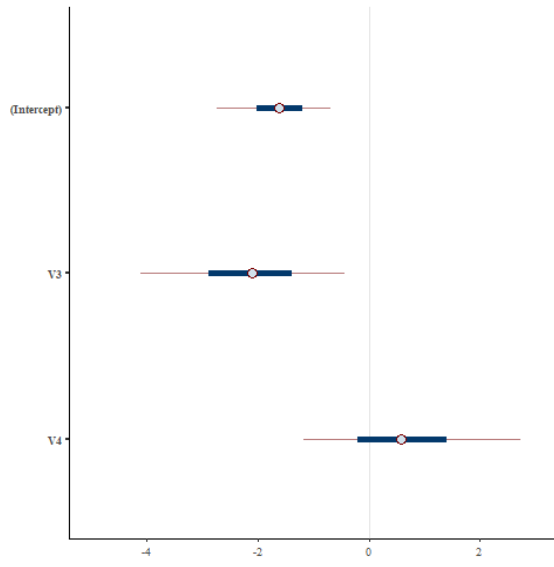
Figure 14: Trace plots - Ring.



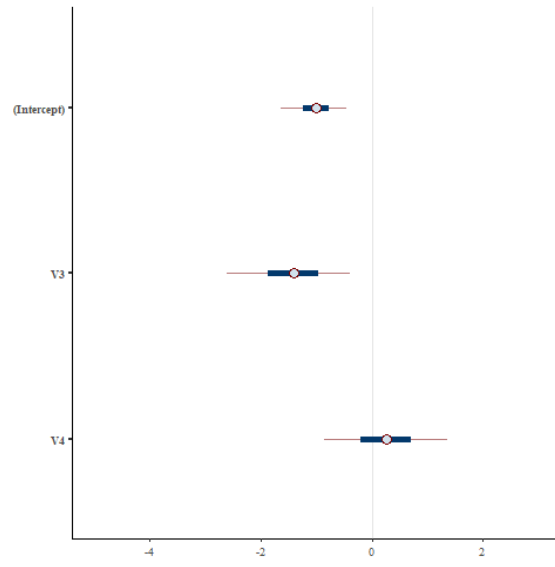
(a)  $\beta \sim N(0, 2.5)$ , logit



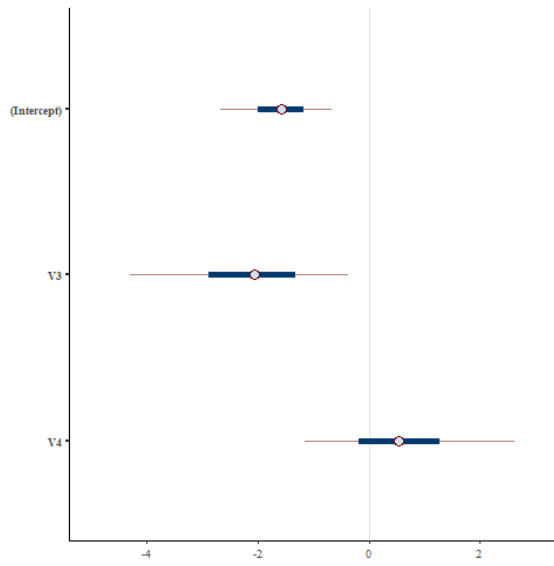
(b)  $\beta \sim N(0, 2.5)$ , probit



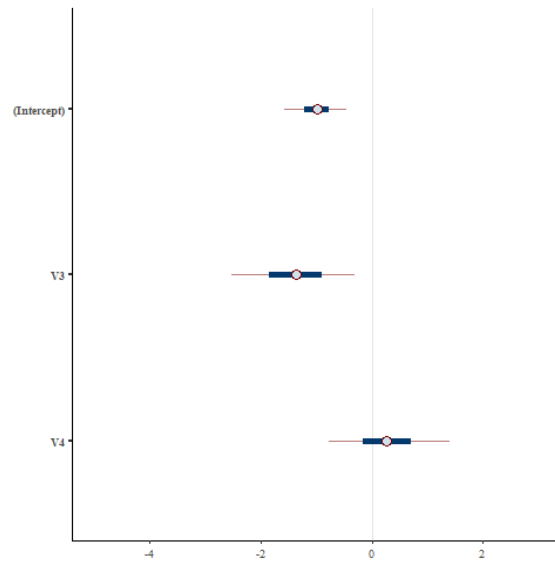
(c)  $\beta \sim t(7, 0, 2.5)$ , logit



(d)  $\beta \sim t(7, 0, 2.5)$ , probit



(e)  $\beta \sim \text{Cauchy}(0, 2.5)$ , logit



(f)  $\beta \sim \text{Cauchy}(0, 2.5)$ , probit

Figura 15: Intervalos de credibilidade - Ring.

	elp_diff	se_diff
ring_fit_11	0.0	0.0
ring_fit_21	-0.2	0.2
ring_fit_31	-0.5	0.3
ring_fit_32	-2.2	1.6
ring_fit_12	-2.4	1.7
ring_fit_22	-2.4	1.8

Tabela 11: Comparação dos modelos usando *loo\_compare()* - Ring.

## Comentários

Um comportamento observado em todas as amostragens é  $Rhat = 1$  e  $mcse = 0$  (ou quase 0), indicando convergência e baixo ruído. Além disso, a grande maioria das amostras apresentam  $n_eff$  alto, indicando baixa correlação na amostragem. Ademais, todos os trace plots apresentam boas misturas, com a cadeia mudando de estado com frequência. Há apenas algumas situações em que a cadeia fica presa em alguns estados durante poucas iterações, como é o caso da figura (8e). Outro comportamento comum foi a alteração das estimativas ao alterar a função de ligação (encurtamento do intervalo de credibilidade e deslocamento da média), conforme as figuras (7), (11), (13), (15). Uma mudança perceptível ao alterar a priori foi observada apenas na variável  $V5$  dos dados Lenses, conforme a figura (9). Por fim, nenhuma das validações cruzadas indicaram preferência significativa por alguma priori, como pode ser observado nas tabelas (2), (9), (6, (4) e (11). Houve apenas uma preferência no ordenamento para a  $t$  em 3 conjuntos de dados. Uma justificativa para isso é que os modelos foram ajustados considerando apenas as prioris normal, student e cauchy, com parâmetros fixos, o ideal seria considerar um conjunto de hiperparâmetros para variar as prioris. Além disso, deveríamos analisar a adequação dos modelos propostos para cada um dos dados, mas exigiria uma análise mais aprofundada.

- (f) Não vou fornecer a prova formal, mas apenas indicar a ideia da prova fornecida em [5]. Primeiro, para a condição necessária: se  $\mathbf{X}_j$  é um separador solitário, então  $\mathbb{E}[\beta_j | \mathbf{y}]$  não existe. O primeiro passo consiste em escrever

$$p(\mathbf{y} | \boldsymbol{\beta}) = \prod_{i \in A_1} f_1(x_{i,j}\beta_j + \mathbf{x}_{i,(-j)}^\top \boldsymbol{\beta}_{(-j)}) \prod_{k \in A_0} f_0(x_{k,j}\beta_j + \mathbf{x}_{k,(-j)}^\top \boldsymbol{\beta}_{(-j)}),$$

onde  $f_1(t)$  é a função sigmóide,  $f_0(t) = 1 - f_1(t)$ , e  $\boldsymbol{\beta}_{(-j)}$  e  $\mathbf{x}_{i,(-j)}$  denotam os vetores sem a  $j$ -ésima entrada. A partir disso, usa-se a suposição de separação para mostrar que

$$p(\mathbf{y} | \boldsymbol{\beta}) \geq \prod_{i \in A_1} f_1(\mathbf{x}_{i,(-j)}^\top \boldsymbol{\beta}_{(-j)}) \prod_{k \in A_0} f_0(\mathbf{x}_{k,(-j)}^\top \boldsymbol{\beta}_{(-j)}).$$

Usando esse fato, mostra-se que

$$\int_0^\infty \beta_j p(\beta_j | \mathbf{y}) d\beta_j \geq \frac{\int_0^\infty \beta_j p(\beta_j) d\beta_j}{p(\mathbf{y})} \left[ \int_{\mathbb{R}^{p-1}} \prod_{i \in A_1} f_1(\mathbf{x}_{i,(-j)}^\top \boldsymbol{\beta}_{(-j)}) \prod_{k \in A_0} f_0(\mathbf{x}_{k,(-j)}^\top \boldsymbol{\beta}_{(-j)}) p(\boldsymbol{\beta}_{(-j)}) d\boldsymbol{\beta}_{(-j)} \right].$$

Por fim, usa-se o fato de que  $\int_0^\infty \beta_j p(\beta_j) d\beta_j = \infty$  (distribuição Cauchy) e como o restante da expressão é finito, conclui-se a prova.

Para a suficiência, o primeiro passo é mostrar que  $\mathbb{E}[\beta_j \mid \mathbf{y}]^+$  é finito. A partir disso, usa-se o fato da não separação para concluir que ou existe

- $i' \in A_1$  tal que  $x_{i',j} < 0$ , ou
- $k' \in A_0$  tal que  $x_{k',j} > 0$ .

Agora, assumindo a primeira condição como verdade, para todo  $\beta_j$ , define-se

$$G(\beta_j) = \{\boldsymbol{\beta}_{(-j)} \in \mathbb{R}^{p-1} : x_{i',(-j)}^\top \boldsymbol{\beta}_{(-j)} < \epsilon \beta_j\},$$

para  $\epsilon = \frac{|x_{i',j}|}{2} = -\frac{x_{i',j}}{2}$ . Então,  $\forall \boldsymbol{\beta}_{(-j)} \in G(\beta_j)$ ,  $\mathbf{x}_{i'}^\top \boldsymbol{\beta} < -\epsilon \beta_j$  e  $f_1(\mathbf{x}_{i'}^\top \boldsymbol{\beta}) < f_1(-\epsilon \beta_j)$ .

Usando isso, conclui-se que

$$\mathbb{E}[\beta_j \mid \mathbf{y}]^+ < \int_0^\infty \beta_j p(\beta_j) \left[ \int_{G(\beta_j)} f_1(-\epsilon \beta_j) p(\boldsymbol{\beta}_{(-j)}) d\boldsymbol{\beta}_{(-j)} + \int_{\mathbb{R}^{p-1} \setminus G(\beta_j)} p(\boldsymbol{\beta}_{(-j)}) d\boldsymbol{\beta}_{(-j)} \right] d\beta_j.$$

A partir disso, a ideia é cotar essas integrais para mostrar a finitude da expressão. Para o caso em que a segunda condição é verdade, é feito apenas um ajuste e a prova segue. Do mesmo modo, para mostrar que a parte negativa  $\mathbb{E}[\beta_j \mid \mathbf{y}]^-$  é finita, usa-se uma ideia bem parecida para encontrar uma cota inferior.

Os códigos e dados utilizados se encontram neste repositório: <https://github.com/EzequielEBS/a2-bayesian-statistics>.

# Referências

- [1] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360 – 1383, 2008.
- [2] P. McCullagh and J. A. Nelder. *Generalized linear models*. Chapman & Hall, 1989.
- [3] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 2003.
- [4] American National Election Studies. 2020 time series study, 2021.
- [5] Joyee Ghosh, Yingbo Li, and Robin Mitra. On the use of cauchy prior distributions for bayesian logistic regression, 2017.